

SHORT COMMUNICATION

# Estimating allelic richness: Effects of sample size and bottlenecks

P. L. LEBERG

Department of Biology, University of Louisiana, Lafayette, LA 70504, USA

## Abstract

Although differences in sampling intensity can bias comparisons of allelic richness ( $A$ ) among populations, investigators often fail to correct estimates of  $A$  for differences in sample size. Methods that standardize  $A$  on the basis of the size of the smallest number of samples in a comparison are preferable to other approaches. Rarefaction and repeated random subsampling provide unbiased estimates of  $A$  with the greatest precision and thus provide greatest statistical power to detect differences in variation. Less promising approaches, in terms of bias or precision, include single random subsampling, eliminating very small samples, using sample size as a covariate or extrapolating estimates obtained from small samples to a larger number of individuals.

*Keywords:* alleles per locus, allelic diversity, bottleneck, randomization, rarefaction, simulation

*Received 19 March 2002; revision received 11 July 2002; accepted 24 July 2002*

## Introduction

There has been considerable interest in using measures of genetic variation to determine whether a population has experienced a demographic bottleneck. There are many different approaches to detecting the genetic signatures of recent bottlenecks (e.g. Richards & Leberg 1996; Luikart *et al.* 1998b). Allelic richness ( $A$ ), also referred to as allelic diversity or mean number of alleles per locus, is one of the most commonly reported measures of genetic variation. Estimates of  $A$  from populations thought to have potentially experienced a bottleneck are compared to estimates from samples collected prior to the putative bottleneck, or to populations with no known history of population reduction (e.g. Stockwell *et al.* 1996; Bouzat *et al.* 1998; Keller *et al.* 2001; Whitehouse & Harley 2001). Both theory and experiments indicate that  $A$  is more sensitive to the effects of short, severe bottlenecks than is heterozygosity, another commonly reported measure of genetic variation (Nei *et al.* 1975; Leberg 1992; Spencer *et al.* 2000). Additionally, it has been argued that allelic richness may reflect more effectively a population's long-term evolutionary potential than would heterozygosity (e.g. Allendorf 1986; Petit *et al.* 1998).

Interpretation of interpopulation differences in  $A$  can be

complicated if based upon unequal numbers of samples. Although Petit *et al.* (1998), Sjogren & Wyoni (1994), Haavie *et al.* (2000), and others have noted the relationship between  $A$  and sample size ( $N$ ), the implications of this relationship are apparently not widely understood. A survey of *Molecular Ecology* (Vol. 9, 2000) yielded 26 research reports that reported estimates of  $A$ , and provides an indication of the scope of the problem. Twenty-two of these studies compared  $A$  among samples of different sizes, but only seven noted that differences in  $N$  might bias estimates of  $A$ . Only five studies attempted to correct for differences in  $N$  or provided an argument for why potential biases in  $A$  were not important. Furthermore, as will be discussed herein, not all of these corrections or justifications had equal merit. *Molecular Ecology* was chosen for this survey only because it publishes many studies examining genetic variation; interpopulation comparisons of  $A$  based on unequal  $N$  are often observed in other journals. This bias has been introduced into my own research (see Leberg 1992).

One reason investigators may continue to compare estimates of  $A$  between samples of different sizes is that they are unaware of the problem. Most treatments of the issue have mentioned the problem only in passing. For example, Sjogren & Wyoni (1994) noted that large sample sizes were needed to detect rare alleles; readers may have failed to make the connection that variation in sampling intensity would bias estimates of  $A$ . Petit *et al.* (1998) and

Correspondence: Paul L. Leberg. Fax: + 337 4825660; E-mail: leberg@louisiana.edu

Haavie *et al.* (2000) discuss the problem directly, but it is not a focus of their studies. Another possibility is that investigators recognize the problem but are unsure of how to correct estimates of  $A$  for different sample sizes. The objective of this study is to discuss several options for addressing the problem associated with comparison of  $A$  between samples of unequal size.

## Methods

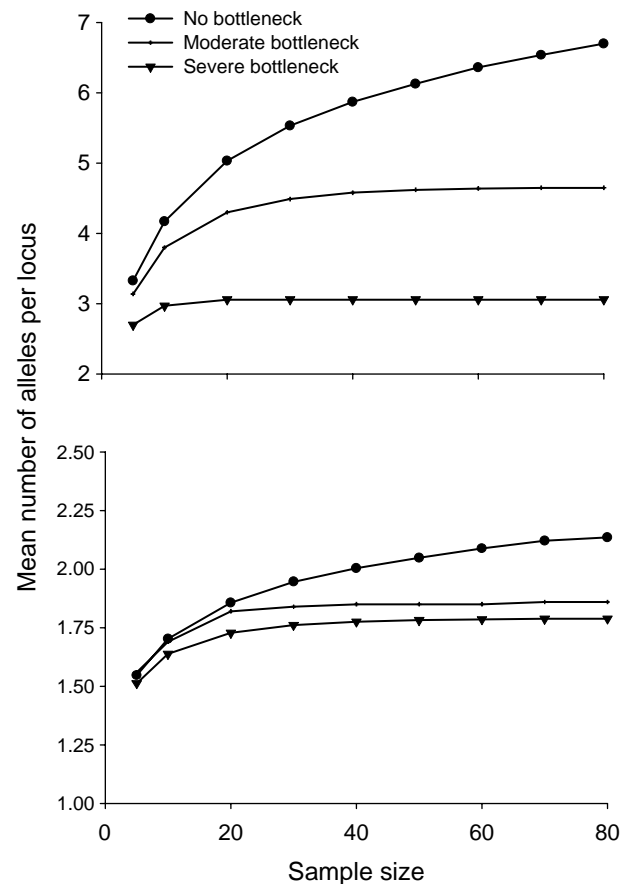
In this evaluation, the average estimated number of alleles per locus ( $\hat{A}$ ) was used as the measure of allelic richness. Computer simulations were used to examine the effects of  $N$  on  $\hat{A}$ . Distributions of alleles with different frequencies were obtained with algorithms provided by Nei *et al.* (1976) for large populations in mutation-drift equilibrium. Two alleles were drawn at random from these distributions to represent each individual in the sample. Sampling bias was examined for loci with two different levels of polymorphism. One set of loci were highly polymorphic with an average  $H_{\text{HWE}} \approx 0.58$  and a mean  $\hat{A} \approx 8.9$  (for a sample of 500 diploid individuals), roughly mimicking the high levels of polymorphism observed for many microsatellite loci. A set of less polymorphic loci had an  $H_{\text{HWE}} \approx 0.18$  and  $\hat{A} \approx 2.5$  ( $N = 500$ ) and was closer to the levels of variation observed in surveys of allozymes and SNPs. For each population sample, 10 distributions of allele frequencies were sampled, representing 10 independent loci. Samples of different numbers of individuals ( $N = 5, 10, 20, 40, 50, 60, 70, 80$ ) were drawn from the distributions to determine allelic richness for each locus;  $\hat{A}$  was estimated as the mean number of alleles at these loci.

To evaluate the effects of bottlenecks on bias of  $\hat{A}$  based on unequal  $N$ , new allele frequency distributions were obtained by simulating bottlenecks of eight or 24 individuals for one generation (referred to as severe and moderate bottlenecks, respectively). These new distributions were created by drawing 16 or 32 alleles, respectively, from the allele frequency distributions obtained for large populations in mutation-drift equilibrium. These new distributions of allele frequencies were subsequently sampled to determine  $\hat{A}$ . Each combination of locus polymorphism, bottleneck history and  $N$  was replicated 1000 times. Results of these simulations were used to help assess different strategies for removing bias from comparisons of based on unequal  $N$ .

## Results and discussion

### Bias due to unequal $N$

Variation in  $N$  can have a large effect on  $\hat{A}$  (Fig. 1). For populations with no history of bottlenecks, samples represented by a small  $N$  will appear to be genetically



**Fig. 1** Allelic richness (mean number of alleles per locus) for different sample sizes based on 1000 simulated populations per treatment combination. Treatment combinations include high (top) and low (bottom) levels of locus polymorphism and exposure or lack of exposure to a simulated bottleneck of eight (severe) or 24 (moderate) individuals for one generation.

depauperate compared to larger samples. The relationship between  $N$  and  $\hat{A}$  is asymptotic, with the greatest effect of differences of  $N$  on  $\hat{A}$  occurring when  $N$  is small. Although no reasonable  $N$  will ensure detection of all alleles, the use of large sample sizes will result in better estimates of  $A$  than the use of small sample sizes.

The level of polymorphism affects the relationship of  $N$  and  $\hat{A}$ . For loci with low polymorphism,  $\hat{A}$  is less biased by differences in  $N$  than are the estimates for more polymorphic loci (Fig. 1). Thus unequal  $N$  will probably bias  $\hat{A}$  for allozymes less than for microsatellites. However, bottlenecks produce smaller differences in  $\hat{A}$  when locus polymorphism is low (Spencer *et al.* 2000), so there is still considerable opportunity for misinterpreting a population's bottleneck history when sample sizes are unequal (Fig. 1).

Because  $A$  decreases as a result of bottlenecks, it is not surprising that the influence of  $N$  on  $\hat{A}$  is reduced following

bottlenecks (Fig. 1). Furthermore, alleles at very low frequencies become less abundant than alleles at intermediate frequencies following bottlenecks (Luikart *et al.* 1998a). Therefore, the effects of  $N$  on  $\hat{A}$  decrease with bottleneck severity (Fig. 1).

#### Potential remedies

Unequal sampling effort is common and can affect  $\hat{A}$ . Various remedies for correcting bias due to unequal  $N$  have been proposed. In discussing potential remedies, singling out specific examples of studies, using problematic adjustments of  $\hat{A}$ , has been avoided. Little is gained from drawing attention to well-intentioned but inadequate attempts to address biases in  $\hat{A}$ , when many studies ignore the problem altogether.

#### Single random reduction in $N$

A simple solution is to reduce all samples to the size of the smallest sample in the comparison. For example, if the smallest  $N$  in the comparison is 10 individuals,  $\hat{A}$  would be estimated from 10 randomly selected individuals from each of the larger samples. This strategy offers a rapid means for obtaining an unbiased estimate of  $A$  for interpopulation comparisons. A disadvantage of this approach is that substantial information is discarded with the individuals not selected for inclusion in  $\hat{A}$ . Based on simulations, the standard deviation (SD) of  $\hat{A}$  is higher than that obtained with other approaches (Table 1); this reduced precision could reduce the power of statistical comparisons of allelic richness among populations.

#### Multiple random reductions of $N$

This approach is similar to making a single random reduction of  $N$ ; however, it is repeated many times. The mean of the multiple samples is the estimate of  $A$ . The SD of this estimate is much smaller than that associated with a single random sample (Table 1). This reduction of variance would increase the statistical power of interpopulation comparisons of  $\hat{A}$ . Thus, multiple subsamples make better use of the information contained in the original sample than does a single subsample. No rule of thumb exists for how many random subsamples should be taken; Haavie *et al.* (2000) used 100 subsamples to obtain  $\hat{A}$ .

It is important to note that this subsampling is not a 'bootstrap' estimate of  $A$ . In a classic bootstrap analysis, a sample of  $N$  individuals (or  $2N$  alleles) would be drawn with replacement from a sample containing  $N$  individuals (Dixon 2001); there is no correction for sample size differences between samples. I am not aware of any software that calculates  $\hat{A}$  based on subsampling, although most spreadsheets and statistical packages can be programmed

**Table 1** Simulations of use of random subsampling and rarefaction to estimate number of alleles for samples of five or 20 individuals from a initial sample size of 80. The means and standard deviations were obtained from 1000 replicate simulations. Estimates were for loci with high polymorphism (mean  $A = 8.9$  for  $N = 500$ ) from populations that had not experienced simulated bottlenecks

Treatment	Mean	SD
Unadjusted estimate for $N = 80$	6.7	1.109
Single random sample of five individuals	3.3	0.453
Multiple (100) subsamples of five individuals	3.3	0.036
Rarefaction to five individuals	3.3	0.042
Single random sample of 20 individuals	5.0	0.770
Multiple (100) subsamples of 20 individuals	5.0	0.081
Rarefaction to 20 individuals	5.0	0.085

to conduct this procedure. I have written a subroutine, SUBSAMPLE, available upon request, to accomplish this task using SAS (1999).

When using random subsampling to standardize  $N$ , it is important to standardize to the smallest  $N$  to be used in a comparison. Reducing the largest samples to the mean  $N$  for all populations, for instance, would reduce but not eliminate bias. Estimates of  $A$  from samples that were smaller than the mean  $N$  would still be biased downward.

#### Rarefaction

This approach uses the frequency distribution of alleles at a locus to estimate the number of alleles that would occur in smaller samples of individuals (Petit *et al.* 1998). Typically, rarefaction is used to standardize  $\hat{A}$  to the smallest  $N$  in a comparison (Petit *et al.* 1998). Estimates of  $\hat{A}$  obtained with rarefaction are similar to those from subsampling (Table 1) and differences in precision, as measured by the SD, are small enough that they would be unlikely to affect the outcome of population comparisons. The genetic analysis programs CONTRIB (Petit *et al.* 1998) and FSTAT Version 2.9.3 (Goudet 2001) incorporate a rarefaction option.

#### Exclusion of estimates based on small $N$

Because the effects of  $N$  on  $\hat{A}$  are greatest for small samples (Fig. 1), one potential solution would be to drop very small samples from any comparison. The effectiveness of this solution clearly depends on the level of variation in the marker system (Fig. 1). For highly polymorphic markers,  $\hat{A}$  has not yet reached an asymptote at 80 individuals. The  $\hat{A}$  obtained when  $N = 80$  is substantially smaller than it would be for  $N = 500$  individuals (6.7 vs. 8.9). However, for the less polymorphic set of loci,  $\hat{A}$  obtained from a sample of 20 individuals was essentially the same as  $\hat{A}$  obtained with sample sizes of 80 or 500. When assaying loci with low

numbers of alleles, eliminating populations represented by only a few individuals may remove most of the bias in  $\hat{A}$  resulting from unequal  $N$ .

Eliminating small samples from comparisons prior to rarefaction or subsampling might be appropriate because with small  $N$ , the difference in  $\hat{A}$  between populations that have and have not experienced a bottleneck is smaller than at larger  $N$ . Take the case where most populations in a comparison are represented by a large number of samples (e.g. between 40 and 60 per population), while one is based on only five samples. The power to detect differences in  $\hat{A}$  would be greatly reduced if  $\hat{A}$  for all of the samples were standardized to  $N = 5$ . Removing the small sample from the comparison would allow the investigator to standardize the remaining estimates of  $\hat{A}$  to  $N = 40$ .

#### *Complete sampling of all or some populations*

If samples were obtained from all of the individuals in a population (Rajora *et al.* 2000), there is no issue of sampling bias as there is no sample. Such complete characterizations of populations are rare unless populations are small. More commonly, complete sampling will be possible for only the smallest populations in any comparison. Complete sampling of small populations with random sampling of large populations does not remove biases in  $\hat{A}$  resulting from unequal sampling efforts. However, if a completely sampled population produced a lower  $\hat{A}$  than a larger  $N$  from an incompletely sampled population, it is safe to conclude that  $\hat{A}$  is lower in the former, even if the difference in  $\hat{A}$  between the two populations is biased.

#### *Extrapolation of $\hat{A}$ from a small to a large sample*

Adjusting  $N$  to the largest number of samples observed, rather than the smallest, would create a more powerful means of detecting differences in  $\hat{A}$  among with the populations because differences in  $\hat{A}$  due to bottleneck history increase with  $N$  (Fig. 1). However, there is no straightforward way to use random sampling to extrapolate estimates of  $A$  to larger samples. Resampling a small sample cannot increase the number of alleles present. It would be difficult to use a relationship between  $\hat{A}$  and  $N$  obtained through subsampling for the purpose of extrapolation, because this relationship is nonlinear.

Another strategy is the use of a theoretical model to predict the number of alleles in a sample. For example, based on a relationship developed by Ewens (1972), the observed number of alleles in a sample has been used to estimate the product of the effective population and mutation rate ( $4N_e\mu$ ), which in turn was used to estimate  $A$  for a larger  $N$ . This strategy is problematic because a large sample size is necessary to estimate the number of alleles accurately. I have conducted simulations (not presented here for

brevity) using the relationship between number of alleles and  $4N_e\mu$  to obtain  $\hat{A}$  for samples that were larger than the original sample. When the size of the original sample is not very different from the size of the larger sample (e.g. 60 and 80), reasonable extrapolations of  $\hat{A}$  are obtained. However, this procedure results in decreased precision of  $\hat{A}$  when a small number of samples, such as five to 10, were used to estimate  $A$  for larger samples of 60–80 individuals.

Other difficulties with extrapolations based on theoretical relationships are associated with the underlying assumptions of the pertinent models. For example, the Ewens's relationship is based on an equilibrium obtained under the infinite alleles model. Because recent bottlenecks will affect the distribution of alleles of different frequencies (Luikart *et al.* 1998a), it is probable that the number of alleles in a population will differ from equilibrium expectations of the infinite alleles model. Similarly, assumptions about the mutational model might not hold for some types of loci. Rarefaction and subsampling require no assumptions concerning equilibrium conditions and mutation models when the objective is to compare  $\hat{A}$  between samples of unequal  $N$ .

#### *Linear adjustment of $A$ by $N$*

Investigators might be tempted to adjust for unequal  $N$  by dividing  $\hat{A}$  by  $N$  or using  $N$  as a covariate in analysis of covariance (ANCOVA). Using the latter approach, the least square means (sometimes called adjusted means) of  $\hat{A}$  should be free of the effects of  $N$ . These adjustments assume a linear relationship between  $N$  and  $\hat{A}$  (Packard & Boardman 1999). However, this assumption is not met because the relationship is clearly nonlinear (Fig. 1), especially when polymorphism is high.

While the problem of nonlinearity might be addressed through transformation, the use of either the ratio  $\hat{A}/N$  or ANCOVA to adjust  $A$  by  $N$  is also based on the assumption that the slopes of relationships between the  $N$  and  $\hat{A}$  are equal (that the lines are parallel) (Packard & Boardman 1999). This assumption is clearly violated, as the relationship between  $\hat{A}$  and  $N$  are not parallel for populations with different histories of bottlenecks (Fig. 1). This interaction could bias estimates of  $\hat{A}$  obtained using  $N$  as a covariate or as divisor; this problem is not correctable with transformation.

In summary, comparisons of  $\hat{A}$  among populations can provide important information on bottleneck history as long as caution is exercised concerning the effects of unequal  $N$ . The best solutions involve collections of large and fairly equal numbers of samples from all populations. When this is not possible, multiple random subsampling and rarefaction both provide reasonable approaches for standardizing  $\hat{A}$  for interpopulation comparisons. Investigators should obtain support for any bottleneck inferred

from  $\hat{A}$  using other lines of genetic evidence (Waples 1989; Luikart *et al.* 1998a, 1998b, 1999).

## Acknowledgements

Funding from NSF Grant DEB-9907567 provided partial support for this research. Helpful comments were provided by F. Ferrara, M. Green, E. Loos, E. Milligan, and three anonymous reviewers. Help with programming was provided by E. Bullard.

## References

- Allendorf FW (1986) Genetic drift and loss of alleles versus heterozygosity. *Zoo Biology*, **5**, 181–190.
- Bouzat JL, Cheng HH, Lewin HA, Westemeier RL, Brawn JD, Paige KN (1998) Genetic evaluation of a demographic bottleneck in the greater prairie chicken. *Conservation Biology*, **12**, 836–843.
- Dixon PM (2001) The bootstrap and the jackknife. In: *Design and Analysis of Ecological Experiments* (eds Scheiner SM, Gurevitch J), pp. 267–288. Oxford University Press, Oxford.
- Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.
- Goudet J (2001) *FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices* Version 2.9.3. Available from [www.unil.ch/izea/software/fstat.html](http://www.unil.ch/izea/software/fstat.html)
- Haavie J, Saetre GP, Moum T (2000) Discrepancies in population differentiation at microsatellites, mitochondrial DNA and plumage colour in the pied flycatcher — inferring evolutionary processes. *Molecular Ecology*, **9**, 1137–1148.
- Keller LF, Jeffery KJ, Arcese P *et al.* (2001) Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers *Proceedings of the Royal Society of London Series B — Biological Sciences*, **268**, 1387–1394.
- Leberg PL (1992) Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution*, **46**, 477–494.
- Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998a) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.
- Luikart G, Cornuet JM, Allendorf FW (1999) Temporal changes in allele frequencies provide estimates of population bottleneck size. *Conservation Biology*, **13**, 523–530.
- Luikart G, Sherwin WB, Steele BM, Allendorf FW (1998b) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology*, **7**, 963–974.
- Nei M, Chakraborty R, Fuerst PA (1976) Infinite allele model with vary mutation rate. *Proceedings of the National Academy of Sciences USA*, **73**, 4164–4168.
- Nei M, Maruyama T, Chakraborty R (1975) Bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.
- Packard GC, Boardman TJ (1999) The use of percentages and size-specific indices to normalize physiological data for variation in body size: wasted time, wasted effort? *Comparative Biochemistry and Physiology A — Molecular and Integrative Physiology*, **122**, 37–44.
- Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.
- Rajora OP, Rahman MH, Buchert GP, Dancik BP (2000) Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Molecular Ecology*, **9**, 339–348.
- Richards C, Leberg PL (1996) Temporal changes in allele frequencies and a population's history of severe bottlenecks. *Conservation Biology*, **10**, 832–839.
- SAS (1999) *SAS Language and Procedures Guide, Version 8*. SAS Institute Inc, Cary, North Carolina.
- Sjogren P, Wyoni PI (1994) Conservation genetics and detection of rare alleles in finite populations. *Conservation Biology*, **8**, 267–270.
- Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology*, **9**, 1517–1528.
- Stockwell CA, Mulvey M, Vinyard GL (1996) Translocations and the preservation of allelic diversity. *Conservation Biology*, **10**, 1133–1141.
- Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, **121**, 379–391.
- Whitehouse AM, Harley EH (2001) Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. *Molecular Ecology*, **10**, 2139–2149.

